PCT/GB 98 / 01893

The Patent Office

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP9 1RH

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed Andrew Gersey

Dated 21 July 1998

Patents Form 1/77

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
E
27 JUN 1997

The Patent Office

30 27 JUN97 E285416 13 000060
P01/7700 25.00 - 9713666.7

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference — SMW/CP5589304

2. Patent application number *(The Patent Office will fill in this part)* — **9713666.7** — 27 JUN 1997

3. Full name, address and postcode of the or of each applicant *(underline all surnames)* — CAMBRIDGE UNIVERSITY TECHNICAL SERVICES LIMITED
THE OLD SCHOOL
CAMBRIDGE
CB2 1TS

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation — GB

4. Title of the invention — BIOSENSOR MATERIALS AND METHODS

5. Name of your agent *(if you have one)* — MEWBURN ELLIS

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)* — YORK HOUSE
23 KINGSWAY
LONDON
WC2B 6HP

Patents ADP number *(if you know it)* — 109006

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month /year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d)) — YES

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	0
Description	36
Claim(s)	0
Abstract	0
Drawing(s)	12

10. If you are also filing any of the following, state

Priority documents	0
Translations of priority documents	0
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	0
Request for preliminary examination and search *(Patents Form 9/77)*	0
Request for substantive examination *(Patents Form 10/77)*	0
Any other documents *(Please specify)*	0

11. I/We request the grant of a patent on the basis of this application.

Signature  Date 26 June 1997

12. Name and daytime telephone number of person to contact in the United Kingdom: DR SIMON M KREMER 0117 926 6411

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed it it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

BIOSENSOR MATERIALS AND METHODS

Technical Field

This invention relates to biosensor materials and methods. Disclosed are methods of generating microorganisms having utility in biosensing, tools which can be generally used in such methods, the microorganisms themselves, and biosensing methods employing such microorganisms.

Background Art

It is frequently desirable to be able to detect small concentrations of analytes in samples, e.g. environmental samples. For instance, to allow more effective management of scarce environmental resources, more efficient and faster methods of assessing environmental pollution are required. At present, molecular-specific monitoring of effluent streams and other environmental matrices requires extensive chemical manipulation of the sample followed by Gas Chromatography (GC) and Mass Spectrometry (MS) analyses. Although these techniques are highly sensitive, sample preparation is necessarily slow and expensive. Consequently, continuous on-site analysis of a variety of environmental matrices cannot be achieved using these methods at reasonable cost.

An alternative method for the determination of phenols and chlorophenols has been proposed using a biosensor based around *Rhodococcus* sp. [see Riedel et al (1993) Appl Microbiol Biotechnol 38: 556-559]. In this method microorganisms are immobilised in an oxygen electrode, and oxygen uptake in response to added substrates was monitored. Although fairly simple and rapid, this method lacks robustness and is not sufficiently sensitive or specific for detecting particular environmental pollutants.

It can thus be seen that the provision of novel materials and methods capable of being used in the field of biosensing would represent a step forward in the art.

Disclosure of Invention

In a first aspect of the invention there is disclosed a method of detecting the presence or absence of an analyte in a sample comprising the steps of:

(a) contacting the sample with a microorganism which expresses a binding agent capable of binding the analyte, wherein the binding of the agent to the analyte causes a detectable signal, and;

(b) observing said microorganism for said detectable signal;

characterised in that said microorganism has been transformed such as to improve the detectability of the signal.

By "observing" is meant ascertaining by any means (directly or indirectly) the presence or absence or the selected signal. which is indicative of the binding event.

By "improve" is meant, inter alia, altering the nature of the signal to one which can be observed more readily or increasing the intensity of the signal (thereby reducing the sensitivity of the means used to observe it).

Thus by using a transformed microorganism, the limitations inherent in wild-type microorganisms such as those used in the prior art may be overcome. In particular more sensitive and robust monitoring methods than those based on natural oxygen uptake can be employed. Methods for generating such transformants are described in further detail below. Such transformed microorganisms are hereinafter referred to as 'biosensors'.

Preferably the analyte is an environmental pollutant, for instance such as may result from industrial or medical applications. Of particular interest is the detection of mono- and poly-aromatic, cyclic, heterocyclic and linear hydrocarbons such as, but not limited to, components of fuels, solvents, propellants, energetics and pesticides (such as may appear on United States EPA Priority Pollutants List and European Community Grey and Black

Lists) and naturally occurring degradation products of these compounds in industrial process media, vapours, effluents, raw water, rivers, ground waters and soils.

Preferably the transformed microorganism is a mycolic acid bacterium. These bacteria form a supra generic group of Gram-positive, non-sporulating bacteria which is comprised of the genera Corynebacterium, Mycobacterium, Nocardia, Rhodococcus, Gordona and Tsukamurella. Members are metabolically diverse and capable of using as sole carbon source (a growth-inducing substrate) a wide range of natural and xenobiotic compounds, including many key environmentally-toxic and/or industrially-important molecules. The mycolic acid bacteria exhibit several structural and physiological features which appear to be specialisations for hydrocarbon degradation, these include a hydrophobic mycolic acid outer cell layer and associated production of extracellular mycolic acid-derived biosurfactants. Most preferably the bacterium is either Rhodococcus or Nocardia.

The detectable signal may be a change in enzyme function(s), metabolic function(s) or gene expression.

Preferably however the signal is ascertained in consequence to an increased expression of a signal protein from a signal gene, more preferably a heterologous signal gene. Many suitable signal proteins (which have a readily detectable activity) are known in the art e.g. ßgalactosidase, which can generate a coloured substrate. Most preferably the activity of the signal protein, or the protein itself, can be estimated photometrically (especially by fluorimetry). For instance green (and red) fluorescent protein, insect luciferase, and photobacterial luciferase. Methods for introducing such genes into appropriate hosts are described in further detail below.

Generally the bound agent/analyte complex will initiate expression of a signal gene which is operably linked to an inducible promoter. The identification of

suitable promoters and/or coding sequences which are operably linked to them (including that of the binding protein) in mycolic acid bacteria, in order to modify said suitable promoters and/or coding sequences to introduce signal genes therein forms one part of the present invention.

As used herein, "promoter" refers to a non-coding region of DNA involved in binding of RNA polymerase and other factors that initiate or modulate transcription from a coding region of DNA whereby an RNA transcript is produced.

An "inducible" promoter requires specific signals in order for it to be turned on or off.

The terms "operatively linked" and "operably linked" refer to the linkage of a promoter to an RNA-encoding DNA sequence, and especially to the ability of the promoter to induce production of RNA transcripts corresponding to the DNA sequence when the promoter or regulatory sequence is recognized by a suitable polymerase. The term means that linked DNA sequences (e.g., promoter(s), structural gene (e.g., reporter gene(s)), terminator sequence(s), are operational or functional, i.e. work for their intended purposes.

As is known to those skilled in the art, the transport and binding proteins (agents) required for the functionality of the inducible promoter, as well as the catabolic enzymes induced by it, will frequently form part the operon containing the promoter, and may thus be identified and isolated along side it using the methods disclosed above. These additional proteins are hereinafter referred to as "operon proteins".

Generally speaking, those skilled in the art are well able to construct vectors and design protocols for recombinant gene expression in common hosts such as E. coli. Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation

sequences, enhancer sequences, marker genes and other sequences as appropriate. For further details see, for example, *Molecular Cloning: a Laboratory Manual:* 2nd edition, Sambrook *et al,* 1989, Cold Spring Harbor Laboratory Press. Many known techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in *Current Protocols in Molecular Biology*, Second Edition, Ausubel et al. eds., John Wiley & Sons, 1992. The disclosures of Sambrook et al. and Ausubel et al. are incorporated herein by reference.

However, the present inventors have recognised that certain methods previously employed in the art which were developed for enteric bacteria such as E. coli may not be the most appropriate for use in mycolic acid bacteria. The mycolic acid layer and associated biosurfactants (which are a defining feature of these bacteria) and thick cell wall confer great resistance to cell lysis protocols known in the art. Similarly, mycolic strains used in the invention may not be laboratory type strains, and may thus exhibit very high levels of nuclease activity. Accordingly, advantageous methods have been developed by the inventors. The methods of identifying, modifying and employing novel inducible promoters and/or coding regions operably linked to them which are appropriate to mycolic acid bacteria are detailed below.

Thus in a second aspect of the invention there is disclosed a method for identifying DNA encoding an inducible promoter which is induced in response to a specific analyte and/or identifying DNA encoding associated operon proteins comprising the steps of:

(a) culturing a source of mycolic acid bacteria in a selective medium containing said specific analyte and being selective for oligotrophic bacteria,

(b) identifying bacteria capable of subsisting on said

medium,

(c) extracting DNA from said bacteria

(d) incorporating said DNA into vectors

(e) cloning said vectors into a suitable host cells

(f) screening the host cells for said inducible promoter and/or proteins in order to identify vectors encoding it.

By "screening" is meant subjected to analysis in order to determine the presence or absence of a particular defined property or constituent. Methods of screening are discussed in more detail below.

As is known to those skilled in the art "oligotrophic bacteria" are bacteria which exhibit a preference for, and persistent slow growth on, low levels of carbon sources. These bacteria are adapted to and predominate in carbon-poor environments (predominantly aquatic habitats where carbon is limiting to μM levels). These bacteria are rarely capable of the very rapid growth as exemplified by the enteric bacterium E. coli, but are by contrast, extremely persistent and metabolically versatile.

Preferably the medium used in the second aspect is a defined minimal medium called hereinafter 'MMRN' which has been developed by the present inventors to screen for oligotrophic mycolic acid-containing bacteria likely to form the basis of the biosensor. This medium preparation is a derivative of von der Osten et al.(1989) but for mycolic acid bacteria sodium citrate and biotin have been shown to be unnecessary. Most importantly, the level of carbon supplement is reduced to oligotrophic levels (<500 μM, more preferably <100 μM). Experiments show that MMRN facilitates simple, selective enrichment for oligotrophic, mycolic acid-containing bacteria as well as providing the basis for testing and characterisation of gene induction. The medium forms a third aspect of the present invention.

DNA may be extracted from the bacteria by any methods known in the art. However, the present inventors have demonstrated that DNA isolation from mycolic acid soil bacteria using standard techniques is inefficient.

Accordingly, several optimised methods of generating total DNA from mycolic bacteria have been developed, as described in more detail below (Examples 3 and 4). These involve bacterial culture in MMRN supplemented with L-glycine, oligotrophic levels of carbon source (80 μM) and removal of biosurfactants by washing in a non-ionic detergent (e.g. Tween 80) prior to a modified alkaline lysis technique. The concept of using a non-ionic detergent at between 0.05 - 0.5 % (preferably 0.1%) in order to facilitate DNA extraction is central to the novel, optimised methods.

"Vector", unless further specified, is defined to include, inter alia, any plasmid DNA, lysogenic phage DNA and/or transposon DNA, in double or single stranded linear or circular form which may or may not be self transmissible or mobilizable, and which can transform a prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g. autonomous replicating plasmid with an origin of replication).

Preferably the host used is E. coli. More preferably it is an E. coli strain carrying one or more of the mcrABC mrr hsdSRM recA and recO mutations, since this is believed to enhance clone recovery when using DNA derived from mycolic acid bacteria. Gene libraries may be readily maintained in these strains.

Preferably the vector used with E. coli further incorporates the cos element (which is well known to those skilled in the art). Because of their capacity and selection for large DNA inserts and efficient transfection rates, cosmid cloning vectors facilitate rapid gene library construction.

Preferably bacteria are further screened, for instance after stage (b), to ensure an absence of catabolic repression. Catabolite repression could seriously compromise the activity of a biosensor since the presence of a more efficient carbon source (such as glucose, succinate or acetate etc.) would lead to repression of hydrocarbon catabolic pathways which forms the basis of the

sensor. To identify strains lacking catabolic repression, the concentrations of an enzyme known to be associated with the catabolic pathway of interest (e.g. catechol 2,3-dioxygenase, which is associated with toluene catabolism) is assessed in (a) selective medium supplemented with the specific analyte, (b) selective medium supplemented with the specific analyte plus a high efficiency carbon source such as glucose (1 mM) and (c) selective medium supplemented with glucose (1 mM) alone. Enzyme activities should be very low or undetectable in the absence of analyte. In the presence of analyte, and glucose plus analyte, the activities should be, within experimental error, very similar.

The present inventors have established that indicates that catabolic genes in mycolic acid bacteria exhibit poor DNA sequence conservation with analogous enzyme genes in Gram negative bacteria. As a result, "reverse genetic" approaches to isolation of novel catabolic pathways are likely to be of limited use when using such published sequence data.

Thus in one embodiment of the second aspect, the host cells are screened for the inducible promoter and/or operon proteins by screening the cells using one or more probes based on the sequence of other promoters and/or operon proteins employed by mycolic acid bacteria in catabolic enzyme production. One example of a source of suitable sequences is the promoter operator region of the R. corallina mac (monoaromatic catabolic operon) the sequence of which has been made available by the present inventors for the first time. This is described in more detail below, and in Example 9. Thus an inducible promoter and/or operon proteins may be identified by providing a nucleic acid molecule having a nucleotide sequence identical to, complementary to, or specifically hybridisable with, the corresponding part of a known, appropriate, mycolic acid bacterial sequence, such as the sequence shown in Fig. 4. Preferably parts of the sequence are used as probes,

preferably of at least 100 nucleotides (but shorter sequences may be employed under high stringency conditions). The use of primers based on the sequence to screen and identify target sequences by PCR is also envisaged.

The identified putative inducible promoter can then be tested to see if it is operational as described in more detail below. Briefly, the putative promoter is provided in a vector upstream of a protein coding sequence (e.g. a reporter gene) at a position in which it is believed to be operatively linked to that coding sequence. A suitable host is transformed with the resulting vector. The presence or absence of the coding sequence expression product, in the presence of the inducing molecule, is determined. For putative transport proteins or catabolic enzymes identified by homology, function can be confirmed as described below.

As an alternative, or in addition to, homology screening, operon proteins which have catabolic enzymic activity can be screened for by their activity. For instance by contacting substrates for the enzymes (the analytes) with the host cells, or extracts therefrom, and observing for degradation products.

This approach can be used when the enzyme concerned may be successfully expressed in the recombinant host cell. For example, the R. corallina mac operon was isolated by screening recombinant E. coli for expression of a catechol 2,3-dioxygenase activity induced in R. corallina when grown on monoaromatic compounds such as toluene. The substrate of the enzyme is catechol, a water soluble 2 hydroxyphenol which does not lyse E. coli.

However, other potential substrates/analytes e.g. toluene are highly toxic to E. coli and may cause its membrane to destabilise leading to cell lysis. Additionally, gene isolation by function is limited to those genes that are expressed in the test bacterium. Because of their evolutionary distance from the mycolic

acid bacteria, established cloning hosts such as E. coli or Gram-positive bacteria such as Bacillus subtilis and Staphylococcus aureus may not effectively recognise mycolic acid bacterial gene regulatory signals and/or may not transport or survive in the presence or xenobiotics per se. Therefore, isolation by acquisition of novel-phenotype cannot easily be accomplished in these hosts.

In addition, when screening for proteins involved in binding or transporting the analyte, or transducing this binding event to the inducible promoter (e.g. transcription factors), it may be necessary to use a host in which other elements of the entire system (i.e. promoter and/or signal gene or catabolic enzymes) are present in order to demonstrate activity.

In order to circumvent these problems, in a most preferred embodiment of the second aspect, vectors comprising the inducible promoter and/or operon proteins are identified by means of a functional screen in a second host. This can avoid the difficulties described above. Preferably this second host is a suitable mycolic acid bacterium.

In order that the vectors can be maintained in the mycolic acid bacteria, they must encode replicons which can function in mycolic acid bacteria. These replicons can be those known in the art (e.g. based on characterised mycolic acid bacterial plasmids pSR1 (Batt et al., 1985). Alternatively the present inventors have provided a novel method of generating supercoiled or circular plasmid DNA from mycolic bacteria, and this method forms one part of the present invention. The diversity of the mycolic acid bacteria means that it is unlikely that a single replicon will be sufficient to construct biosensors in all strains encountered. Novel replicons which can be used either alone or in conjunction (two or more per vector) with other replicons to expand host range therefore provide a useful contribution to the art.

Thus, using the supercoiled/plasmid method of DNA

isolation detailed in Example 4, two previously uncharacterised plasmids pRC100 and pRC158 have been discovered in soil mycolic acid bacteria Rhodococcus corallina and mycolic acid bacterium strain RC158 respectively.

Strain RC158 contains a supercoiled plasmid of approximately 14.57 kb. The plasmid, designated pRC158, contains at least five EcoRI restriction enzyme sites which can be used to digest the plasmid into a specific restriction pattern of five major restriction fragments of 4.3, 3.3, 2.9, 2.2 and 1.6 Kb DNA respectively. An approximately 100 kb plasmid, pRC100, was isolated from R. corallina

Replicons may be identified from novel plasmids by screening fragments obtained therefrom in disabled vectors containing marker proteins (for instance based on pJP7 described below) to see if they can replicate in mycolic acid bacteria.

Novel plasmids isolated using the method, and novel replicon elements isolated from them, form a fourth aspect of the present invention. These, and existing replicons, may be used to construct cloning vectors which replicate in several mycolic acid bacterial strains. Thus it is possible to clone, isolate by function and express specific genes from not only a single "type strain" as is the common practice in molecular biology but also in a variety of mycolic acid bacteria.

It is preferable that the transfer of the vectors comprising the putative inducible promoters and/or operon proteins to the second host (preferably mycolic acid bacteria) from the first host (preferably an established cloning systems such as E. coli) be achieved using bacterial conjugation. Experiments have shown that restriction enzyme activity in newly isolated mycolic acid bacteria effectively limits the efficiency of electroporation of incorrectly methylated plasmid DNA to very low, or undetectable levels. It is known that most

restriction enzymes preferentially act on double stranded DNA substrates. It is known that conjugative DNA transfer, however, involves a single-stranded DNA intermediate and is thus relatively immune to restriction. It is known that the IncPa conjugative plasmid RP4 can transfer its DNA into a wide range of bacteria by conjugation. Accordingly, a series of conjugatively mobilizable mycolic acid bacteria / E. coli shuttle vectors have been constructed by incorporation of a 440 bp region of the RP4 plasmid encoding the origin of transfer (pJP8 figure 1). Experiments have shown that RP4 oriT vectors can be complemented in trans for tra functions allowing conjugative mobilization into a variety of mycolic acid bacteria at high efficiency.

The vectors for use in the most preferred embodiment of second aspect of the invention (i.e. functional screening in a second host), themselves form a fifth aspect of the present invention, such vectors typically comprising:

(a) a replicon for mycolic acid bacteria
(b) a replicon for E. coli
(c) a conjugative origin of transfer
(d) a lambda cos site

An example of such a vector is that termed pJP8 (Figure 5). This comprises (a) pCY104oriV, (b) pBR322 oriV (c) RP4 oriT , and (d)a cos site; however it will be apparent to those skilled in the art that any of these could be substituted for a sequence having similar function, for instance substituting pRC100 or pRC158 minimal replicon sequences for the novel pCY104 replicon. In use such vectors will further comprise a fragment containing the putative inducible promoter and/or operon proteins and optionally a signal protein, such as have been described above.

Thus a gene library can be constructed in a mobilizable cosmid shuttle vector such as pJP8. After in vitro packaging, cosmids can be recovered by adsorption to

E. coli carrying mcrABC mrr hsdSRM recA recO. Given the size of the mycolic acid genome (approximately 4 Mb) a 99% confidence gene library requires approximately 2500 colonies.

To screen for specific functions (either a complete reaction pathway or specific reactions) the packaged cosmids may be adsorbed to E. coli mcrABC mrr hsdSRM recA recO containing an IncP plasmid such as RK2. Since the RK2 plasmid encodes several antibiotic resistance genes, it is modified by random mutagenesis to disable antibiotic resistance genes which are also used as markers in the cosmid vector. From this transformed strain, the mobilizable cosmid shuttle vector may be conjugated into a wide variety of mycolic acid bacteria for functional screening. In any such screen, the choice of mycolic acid bacterial strain will be governed by the known catabolic functions of the strain. Thus entire pathways may be isolated by screening for gain of function. Alternatively, if a particular strain is known to require only one or a few catabolic activities these may be screened for by complementation.

By incorporation of a signal gene adjacent to the cloning site in pJP8 used to construct the gene library, transconjugant mycolic acid bacteria can be screened for inducible expression of a signal protein such as luciferase in the presence of specific molecules. This will rapidly isolate environmentally responsive promoter/operator/regulator elements.

Once identified, by any of the methods of the second aspect of the invention above, the putative inducible promoter and/or operon proteins may be modified by subcloning mutagenesis (typically within E. coli) and screened for enhanced function in mycolic acid bacteria.

The term 'modified' is used to mean a sequence obtainable by introducing changes into the full-length or

part-length sequence, for example substitutions, insertions, and/or deletions. This may be achieved by any appropriate technique, including restriction of the sequence with an endonuclease followed by the insertion of a selected base sequence (using linkers if required) and ligation. Also possible is PCR-mediated mutagenesis using mutant primers.

It may, for instance, be preferable to add in or remove restriction sites in order to facilitate further cloning.

Alternatively, it may be particularly desirable to modify the binding protein/agent in order to modify its specificity and/or affinity for analyte.

Modified sequences according to the present invention may have a sequence at least 70% identical to the sequence of the full or part-length inducible promoter or operon protein as appropriate.. Typically there is 80% or more, 90% or more 95% or more or 98% or more identity between the modified sequence and the authentic sequence. There may be up to five, for example up to ten or up to twenty or more nucleotide deletions, insertions and/or substitutions made to the full-length or part length sequence provided functionality is not totally lost.

Modified promoters and/or operon proteins can be screened for functionality as described above in relation to isolating novel elements.

Nucleic acid encoding the authentic or modified promoter and/or genes encoding the operon proteins (plus such modified proteins themselves) identified or obtained by the method of the second aspect of the invention form a sixth aspect of the invention.

Thus one embodiment of the sixth aspect is the R. corallina mac locus described in Figures 3 and 4 including the promoter and individual operon proteins encoding therein, and modifications thereof.

The authentic or modified promoter identified or

obtained by the method of the second aspect of the invention may be used to inducibly express a heterologous signal protein in a transformed host; this use forms a seventh aspect of the present invention.

In one embodiment of the seventh aspect, there is disclosed a method of transforming a host with a vector encoding the inducible promoter as described above, operably linked to the signal gene (e.g. encoding luciferase).

The vector used in the seventh aspect may remain discrete in the host. Alternatively it may integrate into the genome of the host.

For a potential host (e.g. Corynebacterium) which does not express or generate the other components of the system which may be required to give biosensor function (for instance the operon proteins such as the transport protein to transport analyte into the cell; binding protein to bind analyte thereby inducing the promoter activity; cofactors required for signal protein activity etc.) these components can be added exogenously in order to perform the methods of the first aspect, or can be encoded on the vector used to introduce the inducible promoter. Indeed, as stated above, any transport and binding proteins required for the functionality of the inducible promoter will frequently form part the operon containing the promoter, and may thus be identified and isolated alongside it using the methods disclosed above.

Preferably, however, the host (e.g. a mycolic acid bacterium, either the same or different to that which provided the source of the inducible promoter, but preferably the same) will itself naturally express the other components of the system required to give biosensor function.

Indeed in this latter case, the signal protein gene may be introduced into the host such that it is operably linked to an existing inducible promoter. In this embodiment of the seventh aspect of the invention the

identification and or isolation of the promoter or associated proteins as described above ultimately provides the information required to allow targeting of the gene into this region. Typically this will be achieved by initiating targeted integration using aspects of the sequence forming part of the promoter region or operon.

Direct integration of a signal gene system such as luciferase (e.g. luxAB operon) into an environmentally responsive regulon in a mycolic acid containing bacterium may be more efficient than approaches based on isolation of gene(s) and its/their characterisation followed by construction of the biosensor. This integration can be achieved by transposition or by illegitimate or legitimate recombination between a genetic construct introduced into the cell and the target operon or gene cluster located on either the chromosome or an episomal element. In situations where a gene cluster or operon has been identified as above, by either screening in E. coli or direct functional cloning in a mycolic acid bacterium, site-specific recombination may be used to direct integration of the signal gene(s) (such as luciferase) into the regulon.

Vectors for use in the seventh aspect of the invention, form an eighth aspect of the invention. Such vectors will typically include: (a) the signal gene, plus (b) the inducible promoter, operably linked to the signal gene, or a sequence capable of initiating recombination of the signal gene such that it becomes operably linked with the inducible promoter. Further operon proteins (optionally modified) may also be included in the vector.

Vectors of the eighth aspect of the invention can be readily constructed on the basis of the present disclosure, for instance based on pJP7 (Figure 6) which is described in more detail below. A ninth aspect of the invention is a (biosensor) host transformed with the vectors of the eighth aspect.

In using the transformants of the ninth aspect in the

methods of the first aspect, the signal (such as bacterial luciferase) may be detected extracellularly using a photomultiplier or photodiode or any other photosensitive device. This maintains the cell integrity and thus resistance to environmental shock.

Also embraced within the scope of the present invention are kits for performing the various aspects of the invention. For instance a kit suitable for use in the first aspect may comprise a preparation of the microorganism, plus further means for carrying out the contact or observation steps e.g. buffers, co-factors (e.g. luciferin for addition to luciferase). A kit for performing the second aspect may include any of the following: selective buffer, a non-ionic detergent, any means for carrying out the screening process (e.g. primers, probes, substrates for catabolic enzymes, vectors for transfer into a second host). Kits for performing the seventh aspect may include vectors for generating biosensors plus other means for transforming hosts with them (e.g. buffers etc.).

The invention will now be further described with reference to the following non-limiting Figures and Examples. Other embodiments falling within the scope of the present invention will occur to those skilled in the art in the light of these.

Figures

Figure 1 - shows an agarose gel on which digestions of the novel plasmid pRC100 has been run, as described in Example 5.

Figure 2 - shows an agarose gel on which digestions of the novel plasmid pRC158 has been run, as described in Example 5.

Figure 3 - shows a schematic view of the R. corallina mac operon obtained by functional screening in E. coli, as described in Example 7. The schematic shows location of predicted genes Regulator REG, Transport TRANS, Monooxygenase MONO, Hydroxymuconic semialdehye hydrolase HMSH, Alcohol dehydrogenase ADH, and confirmed gene Catechol 2, 3-dioxygenase CDO. Initiator and terminator codons are shown as half height and full height lines respectively. The location of predicted promoter regions and direction are indicated by arrows. The molecular weights and coordinates of open reading frames are tabulated.

Figure 4 - shows the complete listing of the R. corallina mac operon as described in Example 7. It includes a portion of a putative nitropropane promoter (5' of the regulator).

Figure 5 - shows a schematic diagram of the pJP8 vector of the present invention, as described in Example 8. Plasmid size is about 8.51 kb. pJP8 is a mycolic acid bacterium - E. coli mobilizable cosmid vector. It carries pCY104 replion; is Kanamycin resistant 15 μg/ml mycolic acid bacteria, 50 μg/ml E. coli. It also carries lambda cos site, RP4 oriT site and a multiple cloning site.

Figure 6 - shows a schematic diagram of the pJP7 vector of the present invention, as described in Example 9. Plasmid size is about 10.66 kb. pJP7 is a mobilizable E. coli/Rhodococcus/Nocardia suicide/luciferase integration vector encoding luxAB signal genes, sacB gene and

thiostreppton resistance in Rhodococcus/Nocardia only up to 75 μg/ml (typically 1-10 μg/ml used in selections). The vector is RP4/RK2 mobilizable. By cloning a region of homology into the region upstream of the luxAB cassette, insertion can be targeted.

Examples

Example 1 - A novel medium for oligotrophic screening

"MMRN" is prepared as a multicomponent stock to avoid the production of uncharacterised compounds during autoclaving. A "basic salts" stock is prepared containing 6g/L Na_2HPO_4; 3g/L KH_2PO_4; 1g/L NaCl; 4g/L $(NH_4)_2SO_4$; adjusted to pH 7.4 and made up to 989 mls with distilled water and autoclaved. A "100x A salts" solution is prepared consisting of 20g/L $MgSO_4$; 2000 mg/L $FeSO_4.7H_2O$; 200 mg/L $FeCl_3$; 200 mg/L $MnSO_4.H_2O$ is prepared in distilled water and autoclaved. A "1000x B salts" solution consisting of 500 mg/L $ZnSO_4.7H_2O$; 200 mg/L $CuCl_2.2H_2O$; 200 mg/L $Na_2B_4O_7.10H_2O$; 100 mg/L $(NH_4)_7Mo_6O_{24}.4H_2O$ is prepared in distilled water and autoclaved. To prepare 1 litre of MMRN, sterile solutions of 989 mls basic salts, 10 mls 100 x A salts, 1 ml 1000 x B salts are combined. For solid media, agar is added to 1.4% w/v. Carbon-energy sources are supplemented to 80 μM final concentrations for soluble molecules, or as vapour for insoluble molecules (where their concentration is decided by their individual partition coefficients generally ranging from 3 to 40 μM). Petri plates or liquid cultures are incubated at 28°C to 30°C for up to 72 hours to accumulate sufficient biomass for genetic and biochemical testing.

Example 2 - Isolation of novel strains of mycolic acid containing bacteria from environmental samples using an oligotrophic screen and MMRN

Novel strains are a source of genetic diversity from which biosensors specific for particular xenobiotic compounds can be constructed. To isolate mycolic acid bacteria, for example Rhodococcus / Nocardia, from an environmental matrix such as soil, a rapid isolation

technique is required. Isolation of bacteria from soil using standard laboratory media containing eutrophic levels of carbon preselects for eutrophic bacteria which can grow rapidly under these conditions. Oligotrophic bacteria such as Rhodococcus / Nocardia are rarely successfully isolated on such rich media. This can be carried out using MMRN to specifically enrich for and subsequently purify strains of mycolic acid-containing bacteria which encode catabolic pathways whose expression is induced by a given xenobiotic. This methodology identifies molecules which are not only substrates, but are necessary and sufficient to induce the appropriate catabolic pathway. Soil suspensions from a matrix likely to express a desired phenotype (for instance a site known or believed to have been contaminated with a particular xenobiotic) can be used to inoculate MMRN supplemented with an oligotrophic level of a easily utilised carbon source (50μM). This provides an initial oligotrophic screen. Oligotrophic mycolic acid-containing bacteria are slow growing and may be expected to have formed colonies after 72 hours incubation at 28°C on MMRN paraffin. The incubation temperature appears to be highly selective of soil Nocardioform bacteria; Petri plates incubated at temperatures above 30°C fail to show detectable colonies. Colonies growing on alkanes can be initially screened for Nocardioform phenotype, selecting for crumbling, crenellated colonies, (possibly mucoid on rich media). Gram- and Ziehl-Neelsen-staining tests rapidly identify Gram-positive, mycolic acid-containing bacteria (Place a slide carrying a heat fixed film on a slide carrier over a sink. Flood with carbol fuchsin solution (basic fuchsin 5g; phenol, crystalline, 25g; 95% or absolute ethanol 50 ml; distilled water 500 ml) and heat until steam rises. Leave for 5 minutes, heating occasionally to keep the stain steaming. Wash with distilled water. Flood slide with 20% v/v sulphuric acid; wash off with distilled water, and repeat several times until the film is a faint pink. Finally wash with water.

Treat with 95% v/v ethanol for 2 minutes. Wash with distilled water. Counterstain with 0.2% w/v malachite green. Wash and blot dry. Acid and alcohol fast organisms are red, other organisms are green).

Mycolic acid-containing bacteria may then be screened for specific hydrocarbon-inducible catabolic pathways using MMRN supplemented with the target xenobiotic pollutant. Strains for which the target molecule is growth inducing may then be isolated and used to as a source of genetic regulatory elements for biosensors or as specific biocatalytic functions. Using this protocol mycolic acid containing bacteria have been and may be rapidly identified with novel and useful catabolic properties. This approach is also useful for identification and isolation of mycolic acid containing bacteria with biocatalytic properties.

Example 3 - Method for isolation of total DNA from mycolic acid bacteria

Bacterial strains were inoculated into 10 mls of MMRN supplemented with 500μM glucose 2% w/v L-glycine and incubated at 28°C for 30 to 40 hours. This medium supports relatively rapid growth of mycolic acid bacteria cells. The L-glycine present is misincorporated into peptidoglycan cell wall substantially weakening its resistance to osmotic shock (Katsumata, et al., 1984). Growth on MMRN appears to enhance the uptake of L-glycine and its apparent misincorporation into the cell arabinogalactan. During this growth phase, mycolic acid bacteria produce extensive surfactants which cause the accumulated biomass to clump into pellicles and exhibit a strong surface tension effect. These pellicles, which are highly resistant to lysozyme, may be broken up and the concentration of biosurfactants substantially reduced by washing the cell pellet in several culture volumes of 10 mM Tris pH8.0; 0.1% Tween 80 and finally resuspended in 1ml of 10 mM Tris HCl pH8.0, containing 10 mg/ml lysozyme. The

lysozyme reaction is incubated 60 to 100 minutes at 37°C depending on the strain involved. Lysis is achieved by addition of 2% final (w/v) sodium dodecyl sulphate at 60°C 40 minutes. The nucleic acids are selectively purified from the cellular debris by sequential phenol, phenol: chloroform :isoamyl alcohol (50:48:2 v/v) extractions. Nucleic acids are concentrated by ethanol precipitation in 2 M ammonium acetate. The nucleic acid pellet recovered is washed with 70% ethanol and resuspended in 100µl 10 mM Tris.HCl pH8.0, 1mM EDTA. 2 µl of this sample may be digested using restriction enzymes.

Example 4 - Method to isolate supercoiled/circular plasmid DNA from mycolic acid bacteria

50 mls Rhodococcus was cultured to mid-logarithmic phase in MMRN supplemented with 2% w/v L-glycine, 2% w/v D-glucose.

The cell pellet was washed in 10 mM Tris pH8.0 and 0.1% Tween 80. Resuspend cell pellet in 7.6 ml 6.7% sucrose; 50 mM Tris.HCl; 1 mM EDTA. Add 2 ml 40 mg/ml lysozyme in 10 mM Tris.HCl 1 mM EDTA. Incubate 37°C 15 minutes. Add 970 µl 250 mM EDTA, 50 mM Tris.HCl pH 8.0. Continue incubation for a further 105 minutes 37 °C. Lyse cells by addition of 600 µl 20% SDS 50 mM Tris.HCl, 20 mM EDTA pH 8.0. Incubate 55°C 30 minutes. Shear lysate by vigorous vortexing 30 seconds. Denature DNA by addition of 560 µl freshly prepared 3 M NaOH followed by gently mixing 10 minutes room temperature. Neutralise by addition of 1 ml 2.0 M Tris.HCl pH 7.0 with gentle mixing 10 minutes. Add 2.1 ml 20% SDS 50 mM Tris.HCl, 1 mM EDTA. Mix gently. Add 4.2 ml ice cold 5 M NaCl. Incubate on ice overnight or for several hours at least. Clear the cellular debris by centrifugation at 48000 g 4°C 90 minutes. The supernatant contains the DNA. Decant the supernatant by addition of an equal volume of ice cold isopropanol. Incubate -20°C 30 minutes. Pellet nucleic acids 4°C, 10000g 20 minutes.

Example 5: Novel plasmids and replicons obtained by the method of Example 4

Two multicopy plasmid replicons were isolated using the method of Example 4; pRC158 from strain RC158 and pRC100 from R. corallina.

Both plasmids have been digested with restriction enzymes to produce characteristic restriction patterns (Figures 1 and 2).

Plasmid pRC100, an approximately 100kb supercoiled circular plasmid present in R. corallina was prepared as described in the text. The agarose gel was loaded in lane 1 with Lambda DNA HindIII size markers (23,130 bp; 9,416 bp, 6,557 bp, 4,361 bp, 2,322 bp, 2,027 bp, 564 bp); lanes 2 to 9 inclusive were loaded with pRC100 digested with BamHI (5'GGATCC3'), Bc1I (5'TGATCA3'), Bg1II (5'AGATCT3'), EcoRI (5'GAATTC3'), HindIII (5'AAGCTT3'), KpnI 5'(GGTACC3'), SacI (5'GAGCTC3'), SalI (5'GTCGAC3') restriction endonuclease reactions which were carried out under standard conditions; lane 10 contains undigested (presumable supercoiled) pRC100 DNA; lane 11 pWW110/40121, lane 12 pWW110/4011; lane 13 pWW15/3202; lane 14 pUC18 lane 15 blank. The DNA fragments have been resolved on a 0.8% Agarose Tris-Acetate-EDTA gel. Southern blotting analysis using Gram-negative mono and polyaromatic catechol 2,3-dioxygenases failed to detect significant sequence conservation.

Plasmid pRC158 is a supercoiled plasmid of approximately 14.57 kb. The plasmid was digested with the EcoRI (5'GAATTC3') restriction endonuclease under standard conditions. The DNA fragments have been resolved on a 0.8% Agarose Tris-Acetate-EDTA gel. This pattern is unique and characteristic to pRC158. The plasmid contains at least five EcoRI restriction enzyme sites which can be used to digest the plasmid into a specific restriction pattern of five major restriction fragments of 4.3, 3.3, 2.9, 2.2 and 1.6 Kb DNA respectively.

These plasmids are relatively small, exhibit a high plasmid copy number and are easily isolated from Rhodococcus / Nocardia. Therefore, they possess several characteristics which are suitable for the construction of Rhodococcus / Nocardia cloning vectors.

The DNA sequence of the minimal replicon regions of these plasmids may be determined by screening fragments obtained therefrom in disabled vectors containing marker proteins (for instance based on pJP7 described below) to see if they can replicate in mycolic acid bacteria.

Example 6: Hybridisation screening for novel promoters and/or operon proteins

The test sample (host cells) are contacted with a nucleic acid molecule probe (preferably around 100 nucleotides or more) based on Figure 4 under suitable hybridisation conditions, and any test DNA which hybridises thereto is identified. Such screening is initially carried out under low-stringency conditions, which comprise a temperature of about 37°C or less, a formamide concentration of less than about 50%, and a moderate to low salt (e.g. Standard Saline Citrate ('SSC') = 0.15 M sodium chloride; 0.15 M sodium citrate; pH 7) concentration. Alternatively, a temperature of about 50°C or less and a high salt (e.g. 'SSPE'= 0.180 mM sodium chloride; 9 mM disodium hydrogen phosphate; 9 mM sodium dihydrogen phosphate; 1 mM sodium EDTA; pH 7.4). Preferably the screening is carried out at about 37°C, a form amide concentration of about 20%, and a salt concentration of about 5 X SSC, or a temperature of about 50°C and a salt concentration of about 2 X SSPE. These conditions will allow the identification of sequences which have a substantial degree of similarity with the probe sequence, without requiring the perfect homology for the identification of a stable hybrid. The phrase 'substantial similarity' refers to sequences which share at least 50%

overall sequence identity. Preferably, hybridisation conditions will be selected which allow the identification of sequences having at least 70% sequence identity with the probe, while discriminating against sequences which have a lower level of sequence identity with respect to the probe. After low stringency hybridisation has been used to identify several clones having a substantial degree of similarity with the probe sequence, this subset of clones is then subjected to high stringency hybridisation, so as to identify those clones having a particularly high level of homology with respect to the probe sequences. High stringency conditions comprise a temperature of about 42°C or less, a form amide concentration of less than about 20%, and a low salt (SSC) concentration. Alternatively they may comprise a temperature of about 65°C or less, and a low salt (SSPE) concentration. Preferred conditions for such screening comprise a temperature of about 42°C, a form amide concentration of about 20%, and a salt concentration of about 2 X SSC, or a temperature of about 65°C, and a salt concentration of about 0.2 SSPE.

Example 7 - Cloning aromatic degradative operon from Rhodococcus corallina by functional screening in E coli

To demonstrate the potential mycolic acid bacteria (e.g. Rhodococcus / Nocardia) have as biosensors and biocatalysts as well as to validate the novel genetic tools and approach to cloning of the present invention, a gene cluster or operon associated with aromatic degradation was cloned and isolated from Rhodococcus corallina. This gene cluster / operon appears to be a broad substrate range monoaromatic degradative pathway and has been designated monoaromatic catabolic (mac) gene cluster or operon. R. corallina was isolated from pristine soil in Canada and is an acknowledged Rhodococcus type strain. This strain encodes a broad range of catabolic activities which include toluene, benzoate, phenol, cumine, cyamine. Genetic

induction of the toluene degradative pathway in R. corallina occurs when toluene is supplied as vapour. This is a level of less than 200 ppm in water. Therefore, the sensitivity inherent in the biology of Rhodococcus is very close to those levels expected for biosensors in industrial use. Similar experiments using a naphthalene utilising Rhodococcus which is also supplied as a vapour

Biochemical assays of ring cleavage dioxygenase activities in crude enzyme extracts of R. corallina cells grown on MMRN supplemented with different growth-inducing xenobiotics indicated that the molecular specificity of ring cleavage dioxygenase induction is good. Toluene induced the meta pathway (although some ortho activity was observed) whereas benzoate and phenol exclusively induces the ortho pathway. Xylene, which is very closely related to toluene does not act as a growth inducing substrate. The closely related compounds toluene and benzoate but not xylene induce different ring-cleavage enzymes despite their relatively similar molecular shape. This behaviour and absence of induction with xylene suggests that the receptor for these or metabolites derived from these molecules is sensitive to minor electrostatic changes in their ligand. This strongly asserts that genetically constructed biosensors derived from these receptor molecules and their regulated promoter(s) will exhibit a level of specificity which exceeds that currently available as field test systems.

Since a clear catechol 2,3-dioxygenase activity was induced by toluene, but not by benzoate (indicating that the meta pathway in this strain is specifically induced by toluene), the catechol 2,3-dioxygenase activity can be used as a marker for gene(s), gene cluster(s) or operon(s) involved in its degradation.

The R. corallina catechol 2,3-dioxygenase structural gene was isolated by functional screening of a partial Sau3A restriction enzyme digest-generated gene library in E. coli hsdRMmcrAB for using the commercially available

cosmid cloning vector pWE15 (Wahl et al., 1987).

Because only a single enzyme activity has been used as a functional marker rather than complete acquisition of a phenotype and given the diversity of Rhodococcus / Nocardia metabolism and the genetic incompatibility between mycolic acid bacteria and E. coli it is possible that numerous catechol dioxygenases may exist but only some will be expressed successfully in E. coli. To facilitate expression of cloned DNA irrespective of the presence of an indigenous promoter element, a phage T7 promoter is located adjacent to the pWE15 unique BamHI restriction site into which the rhodococcal DNA was inserted. Phage T7 RNA polymerase (a single polypeptide) is supplied in tráns from pGP1-2Sm. As a functional screen for 2,3-dioxygenase activity, catechol was sprayed onto nutrient agar plates supplemented with 15 μg/ml kanamycin, 50 μg/ml streptomycin, 0.1 mM isopropyl thiogalactoside (IPTG) incubated at 30°C to accumulate biomass. The expression of T7 polymerase is repressed by temperature sensitive phage lambda repressor which is itself expressed from an IPTG induced lacUV5 promoter. Thus incubation at 42°C leads to induction of T7 polymerase expression and so transcription of the pWE15 insert region from the T7 promoter (i.e. one direction of transcript alone).

Using the pGP1-2Sm T7 expression system, two colonies were isolated which encoded the characteristics catechol 2,3-dioxygenase activity from R. corallina. From approximately 3000 colonies of individual primary clones of R. corallina gene library in an E. coli hsdRMmcrAB strain, two colonies were observed to produce a deep yellow colour indicative of catechol 2,3-dioxygenase activity (2-hydroxymuconic semialdehyde) when exogenous catechol was supplied in phosphate buffer (0.1M pH7.4). These clones were designated clone #1 and clone #2. Restriction enzyme mapping of both clone #1 and clone #2 DNA showed that both encode overlapping regions of DNA but were otherwise nonsibling clones; this is compatible with a primary

screening of a cosmid library.

Southern blot analysis of R. corallina total cellular and plasmid DNA confirmed that the isolated catechol 2,3-dioxygenase locus in clones #1 and #2 are contiguous with an approximately 35 kb region R. corallina genomic DNA. The common region to both clones is comprised of seven major EcoRI restriction fragments (8.3, 7.2, 5.2, 4.9, 4.3, 2.4, 2.3 Kb respectively 34.6 kb in total). To confirm the continuity and source of the clone #1 and clone#2 inserts, an aliquot of clone #2 DNA, which contained a slightly longer R. corallina DNA insert than clone #1, was used as a source of DNA to synthesise a radioactive probe to identify homologous DNA restriction fragments present in an EcoRI restriction digest of total cellular R. corallina DNA as well as other bacterial DNA samples. An randomly picked pWE15 clone which did not express catechol 2,3-dioxygenase was chosen as one control (cosmid clone "clone # 4") and E. coli genomic DNA were selected as control DNAs. At the level of accuracy of the gel, the coincidence of the catechol 2,3-dioxygenase clones #1 and clone #2 DNA inserts relative to the genomic R. corallina EcoRI and SmaI restriction maps indicated that no gross deletions or rearrangements had occurred during the cloning. Significantly, there was no evidence for a supercoiled plasmid location for the catechol 2,3-dioxygenase gene indicating that the locus is chromosomally encoded (although pRC100 has been isolated from R. corallina (see Figure 1) this strain does not encode large linear plasmids). To investigate the potential for gene homologs to be identified a Rhodococcus strain RC161 which was isolated from North East England and so is distinct from R. corallina (which also degrades toluene via meta cleavage but was isolated form soil in Canada) was included in the Southern Blot. There were three RC161 EcoRI restriction fragments which exhibited significant DNA sequence conservation with R. corallina sequences in clone #2. The nature of these sequences is

under investigation.

Colony hybridisation to the R. corallina gene library secondary screen using the 2.4 Kb EcoRI restriction fragment of clone #2 as a source of radioactive probe identified four cosmid clones, pWE15#C, pWE15#D, pWE15#B and pWE15#G encoding overlapping regions of the R. corallina chromosome. Thus a region of the R.corallina genome with a contiguous length of approximately 70 kb has been cloned and isolated. These cosmids will provide a source of R. corallina DNA for future experiments.

The 35 Kb region encoded by clones #1 and #2 was mapped using four six base recognition restriction enzymes. An analysis of the map does not indicate inverted DNA map elements which could be consistent with a transposable element. This does not, however, preclude this possibility existing.

The sequence of the operon is described in Example 9 below.

Example 8 - A method for gene isolation from mycolic acid-containing bacteria by functional screening in Corynebacterium glutamicum

A key aspect of this invention is the ability to genetically manipulate a variety of strains or species of mycolic acid-containing bacteria such as Rhodococcus / Nocardia in a simple, effective way so as to clone and isolate gene(s), gene cluster(s) or operon(s) with applications as biosensors or biocatalysis.

The closely related mycolic acid-containing bacterium Corynebacterium glutamicum may be used as a host to express Rhodococcus / Nocardia genetic material. C. glutamicum shares a common cell wall type and probably similar genetic regulation to Rhodococcus / Nocardia but since it has been used extensively for the industrial production of amino acids and nucleotides it has lost or may never had encoded significant xenobiotic catabolic activity. It therefore

represents a good "naïve" host to express Rhodococcus / Nocardia genes.

Restriction enzyme activity in natural isolates of Rhodococcus / Nocardia effectively limits the efficiency of electroporation to very low, or undetectable levels. Most restriction enzymes recognise double stranded DNA exclusively. Because single-stranded DNA is a necessary product of a replication fork, normal restriction enzyme activity in bacterial cells is limited to double stranded DNA substrates. Conjugative DNA transfer in Gram-negative, and most probably between Gram-positive bacteria as well, involves a single-stranded DNA intermediate. Conjugative DNA transfer should thus, generally, be relatively immune to restriction.

The pJP8 plasmid may thus be used to introduce the library in the first host into a suitable mycolic acid bacterium such as corynebacterium or any mycolic acid bacterium which does not encode the desired phenotype.

The pJP8 plasmid is shown in Figure 5. The shuttle vector carries a approximately 400 bp region of the IncP RK2 conjugative plasmid which encodes the origin of transfer. This may be complemented in trans by IncP tra functions maintained on a suitable compatible recombinant plasmid, or as an integrated construct in the host chromosome or by RK2 itself (modified to disrupt its kanamycin resistance gene - a marker used for pJP8). Conjugation involves "effective contact" between the donor and recipient cells, which in this case are E coli encoding complementing tra functions and bearing the mobilizable cosmid vector and a suitable mycolic acid bacterium respectively. Effective contact is the formation of a cytoplasmic bridge between the two cells through which conjugative DNA transfer occurs. Thus donor and recipient cells are grown to mid to late logarithmic phase of growth in Lauria Bertini broth and MMRN supplemented with suitable carbon source at 37°C and 30 C respectively. Donor and recipient cells are washed in prewarmed media and mixed on

a solid support matrix such as Lauria Bertini Agar plate and incubated at 37°C for up to 16 hours. The mating mixture is scraped from the plate and resuspended in 30°C Lauria Bertini broth, from which serial dilutions are prepared and plated on MMRN agar supplemented with drugs to counter select against the donor and recipient and select for the transconjugant mycolic acid bacterium. Commonly, naladixic acid selects against the donor and kanamycin resistance selects against the recipient. Thus, on a plates supplemented with both only the transconjugant may grow. The plates are incubated at 30°C for 40 hours.

Example 9 - DNA sequence of the proximal region of R. corallina mac locus

The DNA sequence of approximately 7 Kb of R. corallina chromosomal DNA surrounding the mac catechol 2,3-dioxygenase has been determined using automated dye terminator sequencing reactions. A schematic of the current state of the data is presented in Figure 3 which shows at least seven genes which have been identified by protein sequence conservation with known protein motif data (nitropropane dioxygenase, a putative regulatory protein orfR, monoaromatic monooxygenase, hydroxymuconic semialdehyde hydrolase, catechol 2,3-dioxygenase, alcohol dehydrogenase).

The sequence of this region in shown in Figure 4.

The predicated gene organisation of the mac associated region is indicative of the presence of possibly two different catabolic gene clusters or operons; one involving the nitropropane dioxygenase the other the mac gene cluster or operon. Such a genetic organisation suggests that a set of divergent promoter elements are located between the predicted regulatory gene orfR and the mac monooxygenase structural gene. Similarly, another promoter could map immediately upstream of the divergent open reading frame which has conservation to nitropropane dioxygenase.

Example 10 - use of the promoter obtained in Example 9

The R. corallina genes identified by sequence conservation or by function are listed in Figure 3. These are potentially useful as catalytic functions in various chemical transformations. The regulatory protein associated with the putative mac operon (possibly encoded by orfR) is involved in the control of transcriptional initiation at its target promoter. This regulatory protein encodes the specificity of the operon and as such is likely to be central to the biosensor function. Subcloning of the regulatory protein and its target promoter could permit novel biosensor activities to be introduced into other Rhodococcus /Nocardia strains. In addition, if this regulatory protein is subjected to mutagenesis, mutants with altered function could be identified (using a luciferase promoter probe driven by the regulated promoter). The regulatory protein has a specific capability to bind its ligand from the environment. It is therefore potentially useful as a protein adsorbent for specific molecules. This could have application in analytical chemistry sample preparation.

An analysis of the 5' region of the predicted genes and the catechol 2,3-dioxygenase reading frame has allowed us to predict the sequence involved in translational initiation. These "ribosome binding sites" can be used as sequence guides or templates for the creation of synthetic oligonucleotides encoding functional Rhodococcus / Nocardia translational initiation sites. Mutagenesis of this region can identify potentially up and down regulating base sequences changes.

The mac promoter region which controls expression of the cloned operon lies between two putative genes (orfR regulatory gene and orfT transport gene). In addition to forming the basis of a biosensor, the promoter and its cognate regulatory system also could be used as an inducible expression system for Rhodococcus / Nocardia and

other mycolic acid-containing bacteria. The sequence of this region encodes the binding sites and regulatory elements or operators involved in control of the mac and possibly other closely linked genes or operons. This region constitutes the first defined sequence for a Rhodococcus / Nocardia promoter region. It can be used as a probe to identify similar sequences within other mycolic acid containing bacteria such as Rhodococcus / Nocardia. This promoter sequence could be used as a region of homology to drive targeted recombination / insertion of signal gene(s) such as Vibrio luciferase.

A vector such as pJP7 (Figure 6) may be used as follows:

The vector is a 'suicide vector' which can be used to drive expression of bacterial luciferase genes in R. corallina. A portion of the mac promoter region (Figure 4) is ligated into the unique pJP7 XbaI restriction site downstream of an E. coli trpA transcriptional terminator. The sacB gene allows counter selection for the integrated plasmid thus selecting for a second cross-over within the plasmid sequences to produce a gene replacement of the wild type gene with an interrupted gene including luciferase. An aspect to this technique is the ability to introduce DNA constructs into the target cell in a hyperrecombinogenic, non-replicating form. Conjugatively mobilised plasmids may represent just such a form in that they may be single-stranded form. Thus the conjugatively mobilised plasmid pJP7 which cannot replicate in mycolic acid bacteria could be used directly to integrate DNA constructs into a wide range of mycolic acid bacterial strains.

Example 11 - Biosensor

The biosensor of the present invention is typically a recombinant mycolic acid containing bacteria which may be Rhodococcus / Nocardia cell. The natural gene-regulatory system which activates expression of catabolic gene(s),

gene cluster(s) or operon(s) in response to the presence of specific class or type of inducing naturally-occurring or xenobiotic carbon substrate(s) has been genetically manipulated to induce the expression of some signal gene(s), such as but not limited to the Vibrio or Photobacterium bacterial luciferase in the presence of the inducer. This manipulation may have involved either incorporation of the signal gene(s) into a chromosomally- or episomally-encoded regulon under the control of a suitable environmentally-regulated promoter, or by direct sub-cloning of the regulated promoter to a rhodococcal / nocardial plasmid or other replicon or episomal element encoding a promoter-less signal gene(s). The genetic manipulation effecting the substitution or supplementation of the natural genes with the signal gene(s)may involve integration of the signal gene(s) gene cluster(s) or operon into the host chromosome, plasmid or other episomal element so as to place it under inducible regulatory control or subcloning of the analyte (particularly hydrocarbon)-responsive promoter to a multicopy plasmid. The integration may involve site-specific recombination, transposition or illegitimate or homology-driven DNA recombination which is another aspect of this invention; however other methods of DNA integration such as the use of polymerase chain reaction (PCR) are not ruled out.

Signal to noise ratio can be readily improved in the recombinant system by enhancing or optimising expression or function of the signal gene, which may be luciferase, by means of improved gene translational signals and/or increasing levels of transcription by either raising transcriptional rates, mRNA stability or gene dosage of the construct (by subcloning to a plasmid or iterative gene integrations into a chromosome, plasmid or other episomal element). Thus, for instance, transcriptional efficiency of the luciferase genes luxAB can be increased by substitution of the Vibrio translational initiation signals with those from the mac operon.

References

von der Osten et al (1989) Biotechnol Letts 11: 11-16.
Wahl et al (1987) Proc Natl Acad Sci 84: 2160-2164.
Katsumata et al (1989) J Bacteriol 159: 306-311.

Fig 1

Figure : Plasmid DNA Isolation from Rhodococcus corallina
Plasmid pRC100, an approximately 100kb supercoiled circular plasmid present in R. corallina was prepared as described in the text. The agarose gel was loaded in lane 1 with Lambda DNA HindIII size markers (23,130 bp; 9,416 bp, 6,557 bp, 4,361 bp, 2,322 bp, 2,027 bp, 564 bp); lanes 2 to 9 inclusinve were loaded with pRC100 digested with BamHI (5'GGATCC3'), BclI (5'TGATCA3'), BglII (5'AGATCT3'), EcoRI (5'GAATTC3'), HindIII (5'AAGCTT3'), KpnI 5'(GGTACC3'), SacI (5'GAGCTC3'), SalI (5'GTCGAC3') restriction endonuclease reactions which were carried out under standard conditions; lane 10 contains undigested (presumable supercoiled) pRC100 DNA; lane 11 pWW110/40121, lane 12 pWW110/4011; lane 13 pWW15/3202; lane 14 pUC18 lane 15 blank. The DNA fragments have been resolved on a 0.8% Agarose Tris-Acetate-EDTA gel. Southern blotting analysis using Gram-negative mono and polyaromatic catechol 2,3-dioxygenases failed to detect significant sequence conservation.

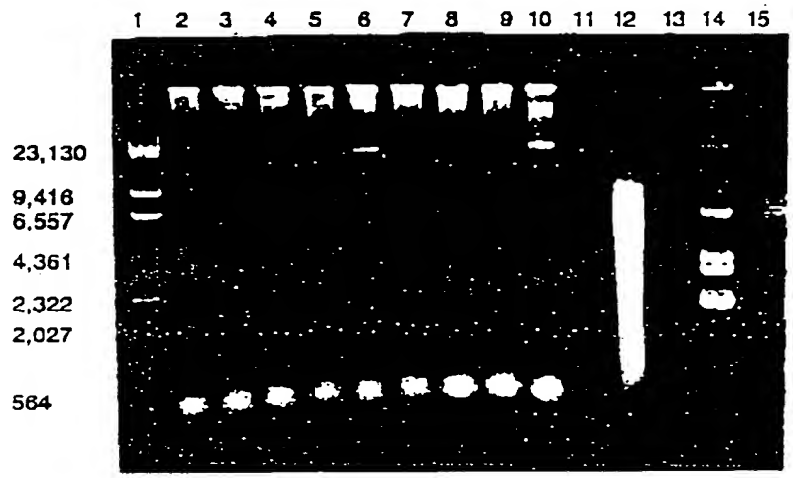


Fig 2

Figure : Plasmid DNA Isolation from Rhodococcus strain RC158 digested with EcoRI

Plasmid pRC158, an approximately 15kb supercoiled circular plasmid present in RC158 was prepared as described in the text and digested with the EcoRI (5'GAATTC3') restriction endonuclease under standard conditions. The DNA fragments have been resolved on a 0.8% Agarose Tris-Acetate-EDTA gel. This pattern is unique and characteristic to pRC158.

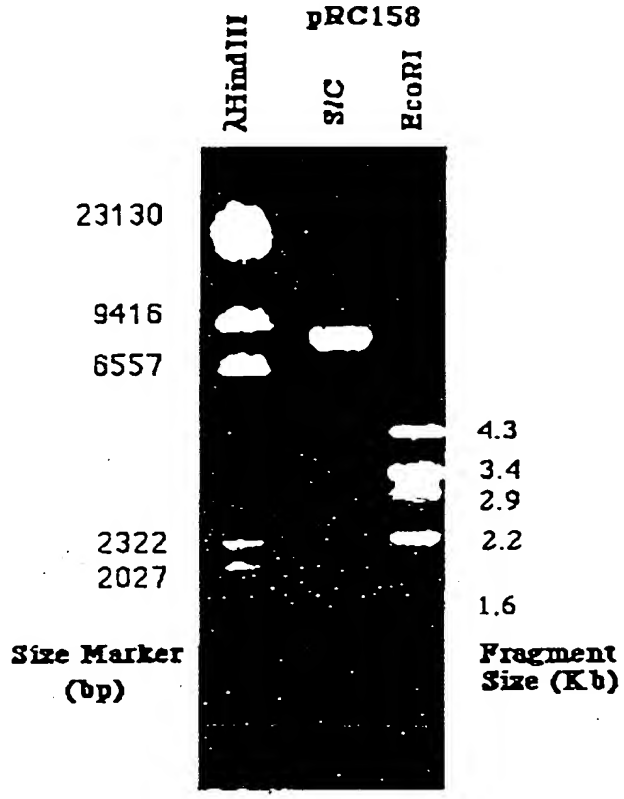
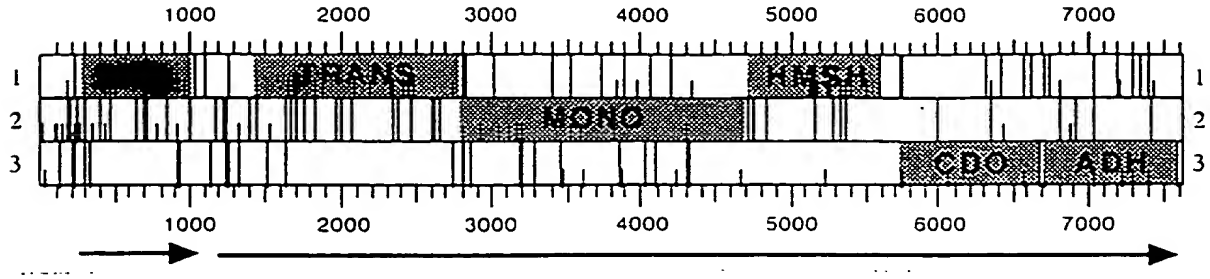


Fig 3



Gene	Initiator Codon	Terminator Codon	Molecular Weight
Regulator	295	1035	27102
Transport	1450	2805	47433
Monooxygenase	2810	4720	69650
Hydroxymuconic semialdehyde hydrolase	4717	5586	32770
Catechol 2,3-dioxygenase	5721	6665	33894
Alcohol dehydrogenase	6711	7580	30586

Figure: Schematic of R. corallina mac operon showing location of predicted genes Regulator REG, Transport TRANS, Monooxygenase MONO, Hydroxymuconic semialdehye hydrolase HMSH, Alcohol dehydrogenase ADH, and confirmed gene Catechol 2,3-dioxygenase CDO. Initiator and terminator codons are shown as half height and full height lines respectively. The location of predicted promoter regions and direction are indicated by arrows. The molecular weights and coordinates of open reading frames are tabulated.

Fig 4a

Figure: Complete listing of R. corallina mac operon encoding predicted regulatory promoter region, regulatory protein, operon promoter region, transport protein, monooxygenase protein, HMSH protein, Catechol 2,3-dioxygenase protein, Alcohol dehydrogenase protein sequences.

```
1/1                                                 31/11
GAA TTC CAT GTT CTT CTC CTT GCA TGT GGC CCG CGT TGC CGA GGG CAC TGC TCG GCC TGT
61/21                                               91/31
CGC CCG CAG AGG GCG CAT GTC CGG GTG CCT GGA TAT GGC GCG TAC GGC GTG CCC TCC GGC
121/41              promoter region                 151/51
GTT AAC CCC GAG GTT GGC CAC GAT GCC CCG GCC ATC AGG TCT GGA ATG CTA GCG TTC CAG
181/61              promoter region                 211/71
ACG AAG GTA ACC CAC AGT GAC TCA CAC CAC AAG TAC TAG AAT GCA AGC TGT TGC GGT GAG
241/81              promoter region                 271/91                      Regulator
CGC CGC GGC ATA AGG GGG AGC CAT GTC CGG GAC GCC GAC GGA AAG CCT GAC TCG ATG ACC
                                                                                M   T
301/101                                             331/111
ACC ACC GAC ACC GGC CCC AAG CCG GGC AGT GAG GCC GCC GCC CTG CTC GCC AAT GTC CGC
T   T   D   T   G   P   K   P   G   S   E   A   A   A   L   L   A   N   V   R
361/121                                             391/131
ACC TCG GGG GCG CGG CTG TCC TCC GCG TTG TAC GAC ATT CTG AAG AAC CGG CTG CTC GAA
T   S   G   A   R   L   S   S   A   L   Y   D   I   L   K   N   R   L   L   E
421/141                                             451/151
GGG CGC TAT GCG GCA GGC GAG AAG ATC GTC GTC GAG TCG ATC CGG CAA GAG TTC GGG GTG
G   R   Y   A   A   G   E   K   I   V   V   E   S   I   R   Q   E   F   G   V
481/161                                             511/171
AGC AAG CAG CCC GTC ATG GAC GCT CTG CGC CGC CTG TCC AGC GAC AAG CTG GTC CAC ATC
S   K   Q   P   V   M   D   A   L   R   R   L   S   S   D   K   L   V   H   I
541/181                                             571/191
GTT CCC CAG GTC GGT TGC GAG GTC GTC TCC TAC GCC CCG CGC GAA GTG GAA GAC TTC TAC
V   P   Q   V   G   C   E   V   V   S   Y   A   P   R   E   V   E   D   F   Y
601/201                                             631/211
ACC CTG TTC GGC GGT TTC GAA GGG ACC ATC GCC GCG GTA GCG GCC TCC CGG CGG ACC GAG
T   L   F   G   G   F   E   G   T   I   A   A   V   A   A   S   R   R   T   E
661/221                                             691/231
GCC CAG TTG CTG GAG CTG GAC CTG ATC TCG GCG CGG GTC GAC GCC CTG ATC ACC TCC CAC
A   Q   L   L   E   L   D   L   I   S   A   R   V   D   A   L   I   T   S   H
721/241                                             751/251
GAC CCG GTG GTC CGC GCC CGC GGG TAC CGC GTG CAC AAC CGG GAG TTC CAT GCG GCC ATC
D   P   V   V   R   A   R   G   Y   R   V   H   N   R   E   F   H   A   A   I
781/261                                             811/271
CAC GCG ATG GCG CAC TCG CGG ATC ATG GAG GAG ACC AGC CAG CGA ATG TGG GAT CTG TGG
H   A   M   A   H   S   R   I   M   E   E   T   S   Q   R   M   W   D   L   S
841/281                                             871/291
GAC TTC TTG ATC AAC ACC ACC GGC ATC ACC AAC CCG CTC TCG AGC GCA CTG CCC GAC CGG
D   F   L   I   N   T   T   G   I   T   N   P   L   S   S   A   L   P   D   R
901/301                                             931/311
CAG CAT GAC CAC CAC GAA ATC ACC GAG GCC ATC CGC AAC CGT GAC GCA GCT GCC GCC CGC
Q   H   D   H   H   E   I   T   E   A   I   R   N   R   D   A   A   A   A   R
961/321                                             991/331
GAG GCC ATG GAA CGC CAC ATC GTC GGC ACC ATC GCA GTA ATC CGC GAC GAA TCC AAC GCC
E   A   M   E   R   H   I   V   G   T   I   A   V   I   R   D   E   S   N   A
1021/341                                            1051/351
CAG CTG CCG AGC TAG ACC CCG ATA CCC GGG CCA TCG ACC GGC TCC GCT ATC GCG CCA CCT
Q   L   P   S   *
1081/361            promoter region                 1111/371
ACG CCG AGG GGG GAC TCT CGG CCG TAG CGC TGC AGA CGA TCC ACC GGC ACC CTC CAC GCT
1141/381            promoter region                 1171/391
```

Fig 46

```
GAC CCC TGT CTC GCC CTA GAG GGC CGG CGC GCC GTC GAT CAC CTT TAC CCT CAT CCA GAG
1201/401              promoter region   1231/411
ACT TGC GTC ACC CTC TAT GCC CGA GTA GCG TCT GAA CTA GAC GTC TAG CAT TCT AGT TGA
Transport
1261/421                                1291/431
GTG CTC CCT CTC GAA GAT TCT CCA GAG AAC CCC TCT CGA ACA TCC CCA GAA GAA AGG AGC
V  L  P  L  E  D  S  P  E  N  P  S  R  T  S  P  E  E  R  S
 C  S  L  S  K  I  L  Q  R  T  P  L  E  H  P  Q  K  K  G  A
  A  P  S  R  R  F  S  R  E  P  L  S  N  I  P  R  R  K  E  R
1321/441                                1351/451
GGC CAT GAC GAC CGC TTC GCA CGC ATC GTC CTT CGG GGC ACG AGC CCA CTT CCG CCC ACA
G  H  D  D  R  F  A  R  I  V  L  R  G  T  S  P  L  P  P  T
1381/461                                1411/471
GAT CGG GGA AGC CCG ACC GTG AGC ACC ACA CCT ACC TCC CCG ACG AAG ACC TCA CCG CTG
D  R  G  S  P  T  V  S  T  T  P  T  S  P  T  K  T  S  P  L
1441/481                                1471/491
CGG GTA GCG ATG GCC AGC TTC ATC GGT ACC ACC GTC GAG TAC TAC GAC TTC TTC ATC TAC
R  V  A  M  A  S  F  I  G  T  T  V  E  Y  Y  D  F  F  I  Y
1501/501                                1531/511
GGC ACC GCG GCC GCG CTG GTA TTC CCT GAG TTG TTC TTC CCG GAT GTC TCG TCC GCG ATC
G  T  A  A  A  L  V  F  P  E  L  F  F  P  D  V  S  S  A  I
1561/521                                1591/531
GGA ATC CTG TTG TCG TTC GCG ACC TTC AGC GTT GGG TTC CTC GCC CGC CCG CTG GGT GGC
G  I  L  L  S  F  A  T  F  S  V  G  F  L  A  R  P  L  G  G
1621/541                                1651/551
ATA GTG TTC GGG CAC TTC GGT GAC CGG GTC GGC CGC AAG CAG ATG CTG GTG ATC TCC CTG
I  V  F  G  H  F  G  D  R  V  G  R  K  Q  M  L  V  I  S  L
1681/561                                1711/571
GTC GGA ATG GGC TCG GCC ACC GTA CTG ATG GGA TTG TTG CCC GGT TAC GCC CAA ATC GGG
V  G  M  G  S  A  T  V  L  M  G  L  L  P  G  Y  A  Q  I  G
1741/581                                1771/591
ATC GCC GCC CCC ATC CTG CTG ACC CTG CTG CGC CTG GTG CAG GGC TTT GCC GTC GGC GGC
I  A  A  P  I  L  L  T  L  L  R  L  V  Q  G  F  A  V  G  G
1801/601                                1831/611
GAG TGG GGT GGA GCC ACC CTG ATG GCC GTC GAG CAC GCC CCC ACC GCG AAG AAG GGC TTT
E  W  G  G  A  T  L  M  A  V  E  H  A  P  T  A  K  K  G  F
1861/621                                1891/631
TTC GGA TCC TTC TCC CAG ATG GGG GCA CCC GCC GGG ACC AGC GTC GCA ACC CTG GCG TTC
F  G  S  F  S  Q  M  G  A  P  A  G  T  S  V  A  T  L  A  F
1921/641                                1951/651
TTC GCG GTC TCC CAA TTG CCC GAC GAG CAG TTC CTG AGT TGG GGC TGG CGA CTG CCG TTC
F  A  V  S  Q  L  P  D  E  Q  F  L  S  W  G  W  R  L  P  F
1981/661                                2011/671
CTG TTC AGC GCG GTG CTG ATC GTG ATC GGG CTG TTC ATT CGC CTG TCC CTG GCC GAA AGC
L  F  S  A  V  L  I  V  I  G  L  F  I  R  L  S  L  A  E  S
2041/681                                2071/691
CCC GAC TTC GCC GAG GTG AAG GCA CAG AGC GCC GTG GTG CGA ATG CCG ATC GCC GAA GCG
P  D  F  A  E  V  K  A  Q  S  A  V  V  R  M  P  I  A  E  A
2101/701                                2131/711
TTC CGC AAG CAC TGG AAG GAA ATT CTC CTC ATC GCG GGC ACC TAC CTG TCC CAA GGA GTG
F  R  K  H  W  K  E  I  L  L  I  A  G  T  Y  L  S  Q  G  V
2161/721                                2191/731
TTC GCC TAT ATC TGC ATG GCC TAC CTC GTC TCC TAC GGC ACC ACC GTC GCG GGG ATC AGC
F  A  Y  I  C  M  A  Y  L  V  S  Y  G  T  T  V  A  G  I  S
2221/741                                2251/751
CGC ACC TTC GCC CTG GCC GGA GTA TTC GTC GCC GGC ATC GTC GCC GTC CTC CTC TAC CTC
R  T  F  A  L  A  G  V  F  V  A  G  I  V  A  V  L  L  Y  L
2281/761                                2311/771
```

Fig 4C

```
GTG TTC GGC GCT CTG TCC GAC ACT TTC GGC CGC AAG ACC ATG TAC CTG CTC GGC GCC GCC
V  F  G  A  L  S  D  T  F  G  R  K  T  M  Y  L  L  G  A  A
2341/781                                    2371/791
GCG ATG GGT GTG GTG ATC GCC CCC GCC TTC GCA CTG ATC AAC ACC GGC AAC CCG TGG CTG
A  M  G  V  V  I  A  P  A  F  A  L  I  N  T  G  N  P  W  L
2401/801                                    2431/811
TTC ATG GCC GCG CAG GTG CTG GTC TTC GGA ATT GCA ATG GCC CCC GCC GCC GGC GTG ACA
F  M  A  A  Q  V  L  V  F  G  I  A  M  A  P  A  A  G  V  T
2461/821                                    2491/831
GGC TCC CTG TTC ACG ATG GTC TTC GAC GCG GAC GTG CGC TAC AGC GGT GTC TCT ATC GGC
G  S  L  F  T  M  V  F  D  A  D  V  R  Y  S  G  V  S  I  G
2521/841                                    2551/851
TAC ACC ATC TCC CAG GTC GCC GGC TCC GCG TTC GCC CCG ACG ATC GCG ACC GCC TTG TAC
Y  T  I  S  Q  V  A  G  S  A  F  A  P  T  I  A  T  A  L  Y
2581/861                                    2611/871
GCC TCC ACC AAC ACC AGC AAC TCG ATC GTG ACC TAC CTG CTG ATC GTC TCG GCC ATC TCG
A  S  T  N  T  S  N  S  I  V  T  Y  L  L  I  V  S  A  I  S
2641/881                                    2671/891
ATC GTC TCG GTG ATC CTG CTG CCC GGC GGC TGG GGG CGC AAG GGC GCT GCG AGC CAG CTC
I  V  S  V  I  L  L  P  G  G  W  G  R  K  G  A  A  S  Q  L
2701/901                                    2731/911
ACT CGC GAC CAG GCC ACC TCC ACA CCG AAA ATG CCT GAC ACC GAA ACA TTT TCG ACT CGG
T  R  D  Q  A  T  S  T  P  K  M  P  D  T  E  T  F  S  T  R
2761/921                                    2791/931                Monooxygenase
ACA GTT CCG GAC ACC GCA GCA TCC CTG CGC GTC CTC GAC AAG TGA AGT GAT GAC AGA CAT
T  V  P  D  T  A  A  S  L  R  V  L  D  K  *  S  D  D  R  H
 Q  F  R  T  P  Q  H  P  C  A  S  S  T  S  E  V  M  T  D  M
2821/941                                    2851/951
GAG TGA CCA CGA CCG CAC CTC CTA CGA CAC CGA CGT CGT GAT CGT CGG CCT CGG CCC CGC
 S  D  H  D  R  T  S  Y  D  T  D  V  V  I  V  G  L  G  P  A
2881/961                                    2911/971
CGG TGG CAC AGC GGC GCT TGC CCT GGC CAG CTA CGG CAT CCG CGT TCA CGC CGT CTC GAT
 G  G  T  A  A  L  A  L  A  S  Y  G  I  R  V  H  A  V  S  M
2941/981                                    2971/991
GTT CCC CTG GGT GGC GAA CTC GCC GCG CGC GCA CAT CAC CAA CCA GCG CGC CGT CGA AGT
 F  P  W  V  A  N  S  P  R  A  H  I  T  N  Q  R  A  V  E  V
3001/1001                                   3031/1011
GCT GCG TGA CCT GGG CGT CGA AGA CGA GGC GCG CAA CTA CGC CAC CCC GTG GGA CCA GAT
 L  R  D  L  G  V  E  D  E  A  R  N  Y  A  T  P  W  D  Q  M
3061/1021                                   3091/1031
GGG CGA CAC GCT GTT CAC CAC GAG CCT GGC CGG CGA GGA GAT CGT CCG GAT GCA GAC CTG
 G  D  T  L  F  T  T  S  L  A  G  E  E  I  V  R  M  Q  T  W
3121/1041                                   3151/1051
GGG TAC GGG CGA TAT CCG CTA CGG GGA CTA CCT GTC CGG AAG CCC CTG CAC GAT GCT CGA
 G  T  G  D  I  R  Y  G  D  Y  L  S  G  S  P  C  T  M  L  D
3181/1061                                   3211/1071
CAT TCC GCA GCC CCT GAT GGA GCC GGT GCT GAT CAA GAA CGC CGC CGA ACG TGG TGC GGT
 I  P  Q  P  L  M  E  P  V  L  I  K  N  A  A  E  R  G  A  V
3241/1081                                   3271/1091
CAT CAG CTT CAA CAC CGA ATA CCT CGA CCA CGC CCA GGA CGA GGA CGG GGT GAC CGT CCG
 I  S  F  N  T  E  Y  L  D  H  A  Q  D  E  D  G  V  T  V  R
3301/1101                                   3331/1111
GTT CCG CGA CGT CCG CTC GGG CAC CGT GTT CAC CCA GCG AGC CCG CTT CCT GCT CGG TTT
 F  R  D  V  R  S  G  T  V  F  T  Q  R  A  R  F  L  L  G  F
3361/1121                                   3391/1131
CGA CGG CGC ACG ATC GAA GAT CGC CGA ACA GAT CGG GCT TCC GTT CGA AGG TGA ACT CGC
 D  G  A  R  S  K  I  A  E  Q  I  G  L  P  F  E  G  E  L  A
3421/1141                                   3451/1151
```

Fig 4d

```
CCG CGC CGG TAC CGC GTA CAT CCT GTT CAA CGC GGA CCT GAG CAA ATA TGT CGC TCA TCG
 R   A   G   T   A   Y   I   L   F   N   A   D   L   S   K   Y   V   A   H   R
3481/1161                                           3511/1171
GCC GAG CAT CTT GCA CTG GAT CGT CAA CTC GAA GGC CGG TTT CGG TGA GAT CGG CAT GGG
 P   S   I   L   H   W   I   V   N   S   K   A   G   F   G   E   I   G   M   G
3541/1181                                           3571/1191
TCT GCT GCG CGC GAT CCG ACC GTG GGA CCA GTG GAT CGC CGG CTG GGG CTT CGA CAT GGC
 L   L   R   A   I   R   P   W   D   Q   W   I   A   G   W   G   F   D   M   A
3601/1201                                           3631/1211
GAA CGG CGA GCC GGA TGT CTC CGA CGA CGT TGT CCT CGA ACA GAT CCG GAC CCT CGT CGG
 N   G   E   P   D   V   S   D   D   V   V   L   E   Q   I   R   T   L   V   G
3661/1221                                           3691/1231
CGA CCC GCA CCT GGA CGT CGA GAT CGT GTC GAG GTC CTT CTG GTA CGT CAA CCG GCA GTG
 D   P   H   L   D   V   E   I   V   S   R   S   F   W   Y   V   N   R   Q   W
3721/1241                                           3751/1251
GGC TGA GCA CTA CCA GTC CGG TCG AGT GTT CTG CGG CGG CGA CGC GGT GCA CCG GCA TCC
 A   E   H   Y   Q   S   G   R   V   F   C   G   G   D   A   V   H   R   H   P
3781/1261                                           3811/1271
GCC GAG CAG CGG GCT GGG CTC GAA CAC GTC CAT GCA GGA CGC GTT CAA CCT GGC ATG GAA
 P   S   S   G   L   G   S   N   T   S   M   Q   D   A   F   N   L   A   W   K
3841/1281                                           3871/1291
GAT CGC GTT CGT CGT GAA GGG GTA TGC AGG ACC GGG TCT GCT CGA GTC CTA CTC TCC TGA
 I   A   F   V   V   K   G   Y   A   G   P   G   L   L   E   S   Y   S   P   E
3901/1301                                           3931/1311
GCG TGT TCC GGT CGG CAA ACA GAT CGT CGC TCG CGC CAA CCA GTC CCG CAA GGA CTA CGC
 R   V   P   V   G   K   Q   I   V   A   R   A   N   Q   S   R   K   D   Y   A
3961/1321                                           3991/1331
CGG GCT GCG CGA ATG GTT CGA TCA CGA GAG CGA CGA CCC GGT CGC CGC CGG CCT GGC AAA
 G   L   R   E   W   F   D   H   E   S   D   D   P   V   A   A   G   L   A   K
4021/1341                                           4051/1351
GTT GAA GGA ACC CTC GTC CGA AGG TGT TGC TCT GCG TGA GCG GCT GTA CGA GGC GCT GGA
 L   K   E   P   S   S   E   G   V   A   L   R   E   R   L   Y   E   A   L   E
4081/1361                                           4111/1371
GGT GAA GAA CGC CGA ATT CAA CGC CCA GGG CGT CGA ACT CAA CCA GCG CTA CAC CTC GTC
 V   K   N   A   E   F   N   A   Q   G   V   E   L   N   Q   R   Y   T   S   S
4141/1381                                           4171/1391
CGC GGT CGT TCC CGA CCC CGA GGC GGG CGA GGA AGT GTG GGT GCG CGA TCG TGA GCT GTA
 A   V   V   P   D   P   E   A   G   E   E   V   W   V   R   D   R   E   L   Y
4201/1401                                           4231/1411
CCT GCA GGC CAC CAC CCG GCC GGG CGC GAA GCT GCC GCA TGC GTG GCT GGT CGG CGC CGA
 L   Q   A   T   T   R   P   G   A   K   L   P   H   A   W   L   V   G   A   D
4261/1421                                           4291/1431
CGG AAC CCG CAT CTC CAC CCT CGA CGT CAC CGG CAA GGG AAT GAT GAC CCT GCT GAC CGG
 G   T   R   I   S   T   L   D   V   T   G   K   G   M   M   T   L   L   T   G
4321/1441                                           4351/1451
ACT CGG CGG CCA GGC ATG GAA GCG TGC CGC CGC CAA ACT CGA CCT GCC GTT CCT GCG GAC
 L   G   G   Q   A   W   K   R   A   A   A   K   L   D   L   P   F   L   R   T
4381/1461                                           4411/1471
CGT CGT TGT CGG CGA ACC CGG CAC CAT CGA CCC TTA CGG ATA CTG GCG GCG GGT CCG CGA
 V   V   V   G   E   P   G   T   I   D   P   Y   G   Y   W   R   R   V   R   D
4441/1481                                           4471/1491
CAT CGA CGA GGC CGG CGC CCT GCT CGT GCG GCC CGA CGG CTA CGT CGC GTG GCG ACA CAG
 I   D   E   A   G   A   L   L   V   R   P   D   G   Y   V   A   W   R   H   S
4501/1501                                           4531/1511
TGC TCC GGT CTG GGA CGA CAC CGA AGC GCT CAC CAG CCT CGA GAA CGC TCT CAC CGC GGT
 A   P   V   W   D   D   T   E   A   L   T   S   L   E   N   A   L   T   A   V
4561/1521                                           4591/1531
CCT CGA CCA CTC GGC CAG CGA CAA CGG GAA CCC GAG CGG CAC AAA CGA GCC GCA GTA CAG
```

Fig 4e

```
L  D  H  S  A  S  D  N  G  N  P  S  G  T  N  E  P  Q  Y  S
4621/1541                                   4651/1551
CAC CCG GGC CGT GCC GAT CGT CGT TCC GCA CGT TAC CGC CGA GGA TGC AGC ACC AGC TTC
T  R  A  V  P  I  V  V  P  H  V  T  A  E  D  A  A  P  A  S
                                          HMSH
4681/1561                                   4711/1571
CGC CAC CCG CAC CAC CAC AGT CGA GGG AGA GAA CCG ATG ACC CGT CCT TAC ACC AGC GTC
R  H  P  H  H  H  S  R  G  R  E  P  M  T  R  P  Y  T  S  V
A  T  R  T  T  T  V  E  G  E  N  R  *  P  V  L  T  P  A  S
4741/1581                                   4771/1591
TGG GAC GAC CTG AAC CAG GTC GAG TTC AGC CAG GGA TTC ATC CAG GCC GGC CCC TAC CGG
W  D  D  L  N  Q  V  E  F  S  Q  G  F  I  Q  A  G  P  Y  R
4801/1601                                   4831/1611
ACC CGA TAC CTG CAC GCC GGC GAT TCG TCC AAG CCC ACG CTG ATC CTG CTG CAC GGC ATC
T  R  Y  L  H  A  G  D  S  S  K  P  T  L  I  L  L  H  G  I
4861/1621                                   4891/1631
ACC GGC CAC GCC GAG GCG TAC GTG CGC AAT CTG CGC TCG CAT TCC GAG CAC TTC AAC GTC
T  G  H  A  E  A  Y  V  R  N  L  R  S  H  S  E  H  F  N  V
4921/1641                                   4951/1651
TGG GCA ATC GAC TTC ATC GGC CAC GGC TAT TCG ACC AAG CCC GAC CAC CCG CTC GAG ATC
W  A  I  D  F  I  G  H  G  Y  S  T  K  P  D  H  P  L  E  I
4981/1661                                   5011/1671
AAG CAC TAC ATC GAC CAC GTG CTG CAG TTG CTG GAC GCC ATC GGC GTC GAG AAG GCC TCG
K  H  Y  I  D  H  V  L  Q  L  L  D  A  I  G  V  E  K  A  S
5041/1681                                   5071/1691
TTT TCC GGG GAG TCT CTC GGC GGT TGG GTC ACC GCC CAG TTC GCG CAC GAC CAT CCC GAG
F  S  G  E  S  L  G  G  W  V  T  A  Q  F  A  H  D  H  P  E
5101/1701                                   5131/1711
AAG GTC GAC CGG ATC GTG CTC AAC ACC ATG GGC GGC ACC ATG GCC AAC CCT CAG GTG ATG
K  V  D  R  I  V  L  N  T  M  G  G  T  M  A  N  P  Q  V  M
5161/1721                                   5191/1731
GAA CGT CTC TAT ACC CTG TCG ATG GAA GCG GCG AAG GAC CCG AGC TGG GAA CGC GTC AAA
E  R  L  Y  T  L  S  M  E  A  A  K  D  P  S  W  E  R  V  K
5221/1741                                   5251/1751
GCA CGC CTC GAA TGG CTC ATG GCC GAC CCG ACC ATG GTC ACC GAC GAC CTG ATC CGC ACC
A  R  L  E  W  L  M  A  D  P  T  M  V  T  D  D  L  I  R  T
5281/1761                                   5311/1771
CGC CAG GCC ATC TTC CAG CAG CCG GAT TGG CTC AAG GCC TGC GAG ATG AAC ATG GCA CTG
R  Q  A  I  F  Q  Q  P  D  W  L  K  A  C  E  M  N  M  A  L
5341/1781                                   5371/1791
CAG GAC CTC GAA ACC CGC AAG CGG AAC ATG ATC ACC GAC GCC ACT CTC AAC GGC ATC ACG
Q  D  L  E  T  R  K  R  N  M  I  T  D  A  T  L  N  G  I  T
5401/1801                                   5431/1811
GTG CCC GCG ATG GTG CTG TGG ACC ACC AAG GAC CCC TCC GGT CCG GTC GAC GAA GCC AAG
V  P  A  M  V  L  W  T  T  K  D  P  S  G  P  V  D  E  A  K
5461/1821                                   5491/1831
CGC ATC GCC TCC CAC ATC CCG GGC GCC AAG CTG GCC ATC ATG GAG AAC TGT GGC CAC TGG
R  I  A  S  H  I  P  G  A  K  L  A  I  M  E  N  C  G  H  W
5521/1841                                   5551/1851
CCC CAG TAC GAG GAC CCC GAG ACC TTC AAC AAG CTG CAT CTG GAC TTC CTC CTC GGT CGC
P  Q  Y  E  D  P  E  T  F  N  K  L  H  L  D  F  L  L  G  R
5581/1861                                   5611/1871
AGC TGA CAC AGA CCC CGG CCG GTG CCG CCA ACC CCT GCA ACC CGG GCG GCA CCG GCC GGA
S  *  H  R  P  R  P  V  P  P  T  P  A  T  R  A  A  P  A  G
 A  D  T  D  P  G  R  C  R  Q  P  L  Q  P  G  R  H  R  P  D
   L  T  Q  T  P  A  G  A  A  N  P  C  N  P  G  G  T  G  R  I
5641/1881                                   5671/1891
TCT CAC TTA CCC GAC CTA TTG CGC TCT CGT CCG GAC CCC CGG AGA GAA AGC GCC GAA GCA
```

Fig 4f

Catechol 2,3-dioxygenase

5701/1901 5731/1911
GCA GCA AGG AGA CCG CCG CGA TGC CTG TAG CGC TGT GCG CGA TGT CGC ACT CCC CCC TGA
S K E T A A M P V A L C A M S H S P L M

5761/1921 5791/1931
TGG GAC GCA ACG ACC CCG AAC AGG AAG TCA TCG ACG CCG TCG ACG CCG CAT TCG ACC ACG
G R N D P E Q E V I D A V D A A F D H A

5821/1941 5851/1951
CGC GCC GGT TCG TCG CCG ACT TCG CCC CCG ATC TCA TCG TCA TCT TCG CCC CCG ACC ACT
R R F V A D F A P D L I V I F A P D H Y

5881/1961 5911/1971
ACA ACG GCG TCT TCT ACG ACC TGC TGC CGC CGT TCT GTA TCG GTG CCG CCG CGC AGT CCG
N G V F Y D L L P P F C I G A A A Q S V

5941/1981 5971/1991
TCG GCG ACT ACG GCA CCG AAG CCG GCC CTC TCG ACG TCG ACC GTG ACG CCG CCT ACG CAG
G D Y G T E A G P L D V D R D A A Y A V

6001/2001 6031/2011
TCG CCC GCG ACG TCC TCG ACA GCG GCA TCG ACG TCG CAT TCT CCG AAC GCA TGC ACG TCG
A R D V L D S G I D V A F S E R M H V D

6061/2021 6091/2031
ACC ACG GAT TCG CCC AAG CAC TCC AAT TGC TGG TCG GAT CGA TCA CCG CCG TGC CGA CCG
H G F A Q A L Q L L V G S I T A V P T V

6121/2041 6151/2051
TGC CGA TCT TCA TCA ATT CGG TCG CCG AAC CGC TCG GCC CGG TCA GCC GGG TAC GGC TGC
P I F I N S V A E P L G P V S R V R L L

6181/2061 6211/2071
TCG GCG AGG CGG TCG GGC GGG CCG CTG CCA AGC TGG ACA AGC GTG TGC TGT TCG TCG GAT
G E A V G R A A A K L D K R V L F V G S

6241/2081 6271/2091
CCG GCG GCC TGT CCC ACG ACC CGC CGG TCC CGC AGT TCG CCA CCG CGC CAG AGG AAG TGC
G G L S H D P P V P Q F A T A P E E V R

6301/2101 6331/2111
GCG AGC GGT TGA TCG ACG GCC GCA ATC CCA GTG CCG CCG AAC GTG ATG CCC GCG AAC AGC
E R L I D G R N P S A A E R D A R E Q R

6361/2121 6391/2131
GCG TCA TCA CCG CCG GGC GGG ACT TCG CCG CCG GCA CCG CCG CCA TCC AGC CAC TGA ACC
V I T A G R D F A A G T A A I Q P L N P

6421/2141 6451/2151
CCG AAT GGG ACC GGC ACC TGC TCG ACG TCC TCG CCT CCG GCG ACC TCG AGC AGA TCG ACG
E W D R H L L D V L A S G D L E Q I D A

6481/2161 6511/2171
CGT GGA CCA ACG ACT GGT TCG TCG AAC AGG CCG GAC ACT CCT CCC ACG AAG TGC GCA CCT
W T N D W F V E Q A G H S S H E V R T W

6541/2181 6571/2191
GGA TCG CCG CGT ACG CGG CAA TGA GCG CCG CCG GGA AGT ACC GCG TCA CCT CGA CCT TCT
I A A Y A A M S A A G K Y R V T S T F Y

6601/2201 6631/2211
ACC GCG AAA TCC ACG AGT GGA TAG CAG GAT TCG GGA TTA CTA CCG CCG TCG CCG TCG ACG
R E I H E W I A G F G I T T A V A V D E

6661/2221 6691/2231 Alcohol dehydrogenase
AAT AGA CCC CGC CGC TCC CGC CCC GCA GTC CCA ACG AAG GGT GGC CCC GGA TGA CCT CCG
* T P P L P P R S P N E G W P R M T S V

6721/2241 6751/2251
TCC GCC CGT GCT CGC CGT CGG TGA ACG CGG GCT GGT CGG TGG GCA GGA AGA CCT CAT CGC
R P C S P S V N A G W S V G R K T S S P

6781/2261 6811/2271
CGA CAT CGC CCT CGA CCT CGC AGC TCG TCA GTA GGA ATG CGC ACG GGC CGA CGA GTC GCG

Fig 4g

T S P S T S Q L V S R N A H G P T S R A

6841/2281 6871/2291

CTG GTC ACC GGG GCC AGC CGC GGC ATC GGG GCG GCC ATC GCA GAT GCG GTG GCC GCC TCC

G H R G Q P R H R G G H R R C G G R L R

6901/2301 6931/2311

GGT GCC GCC GTA ATC GTC CAC TAC GGA TCC GAT CGG ACG GCC GCC GCT GCG GTG TCG ACG

C R R N R P L R I R S D G R R C G V D G

6961/2321 6991/2331

GCA TCA CGG CTG CCG GGG GCC TCG CGG CTG CGG TCC AGG CCG ACC TGT CCC GAC CCG AGG

I T A A G G L A A A V Q A D L S R P E G

7021/2341 7051/2351

GGC CTG AAG AGC TGA TGC GGG AGT TCG ACT CCG CGC TCG ACG GTC TCG GGC TCG ACC GAG

P E E L M R E E D S A L D G L G L D R G

7081/2361 7111/2371

GGC TCG ACA TCC TCG TCA ACA ACG CCG GAA TCA GTC GGC GCG GAG CGC TCG AGC GCG TCA

L D I L V N N A G I S R R G A L E R V T

7141/2381 7171/2391

CTG TCG AGG ATT TCG ACC GTC TGG TCG CAC TCA ACC AGC GCG CCC CGT TCT TCG TGA CTC

V E D E D R L V A L N Q R A P E E V T R

7201/2401 7231/2411

GGC ATG CCC TGC CCC GGA TGC ACG ACG GCG GTC GCA TCG TCA ACA TTT CCT CCG GAT CCG

H A L P R M H D G G R I V N I S S G S A

7261/2421 7291/2431

CCC GCT ACG CCA GAC CCG ACG TCA TCA GCT ACG CCA TGA CCA AGG GGG CGA TCG AGG TGC

R Y A R P D V I S Y A M T K G A I E V L

7321/2441 7351/2451

TCA CCC GCG CCC TCG CCG TAG ACG TCG GCG AAC GAG GCA TCA CCG CCA ACG CCG TGG CGC

T R A L A V D V G E R G I T A N A V A P

7381/2461 7411/2471

CGG CCG CGC TCG ATA CCG ACA TGA ACG CGC ACT GGC TTC GCG GTG ACG ACC ATG CCC GCA

A A L D T D M N A H W L R G D D H A R T

7441/2481 7471/2491

CCA CCG CCG CGT CCA CCA CTG CAC TGC GAA AAC TCG CCA CCG CGG AGG ACA TCG CCG CGA

T A A S T T A L R K L A T A E D I A A I

7501/2501 7531/2511

TCG TGG CCT TCC TCG TCA GCG CCG CCG CCG GTG CGA TCA CCG GGC AGG TCA TCG ACG CCA

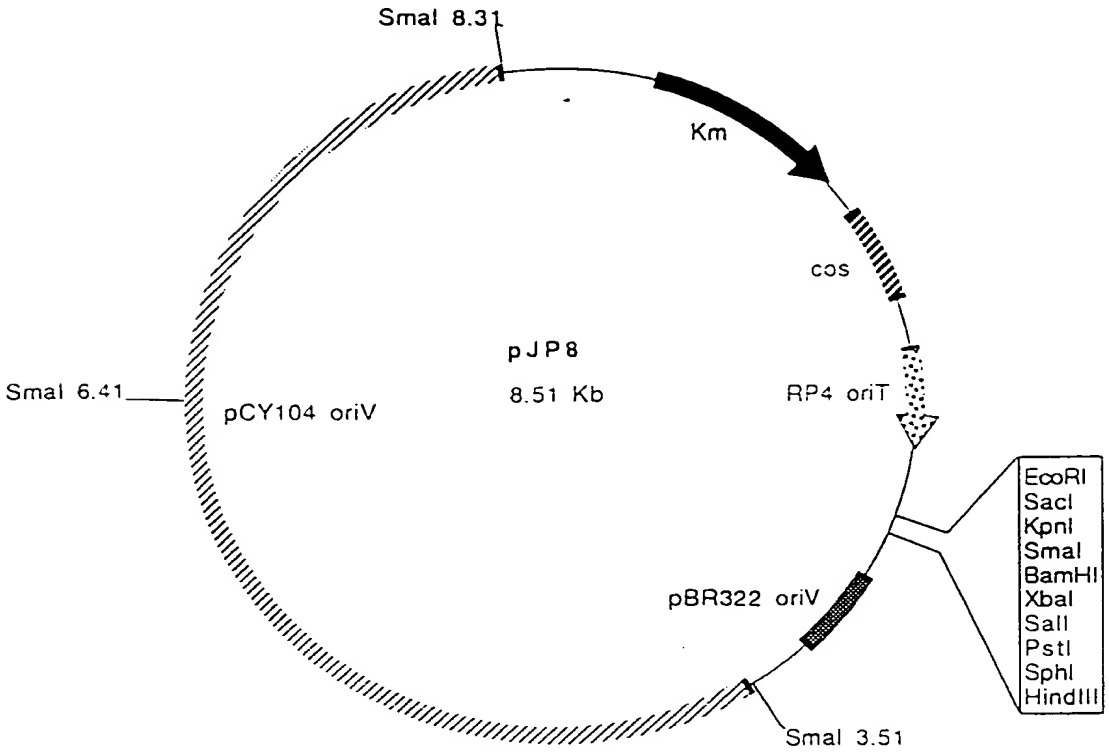
V A F L V S A A A G A I T G Q V I D A T

7561/2521 7591/2531

CCA ACG GCA ACC GGC TCT AAC CAG AAC TTA CCC GGT CCC

N G N R L * P E L T R S

Fig. 5



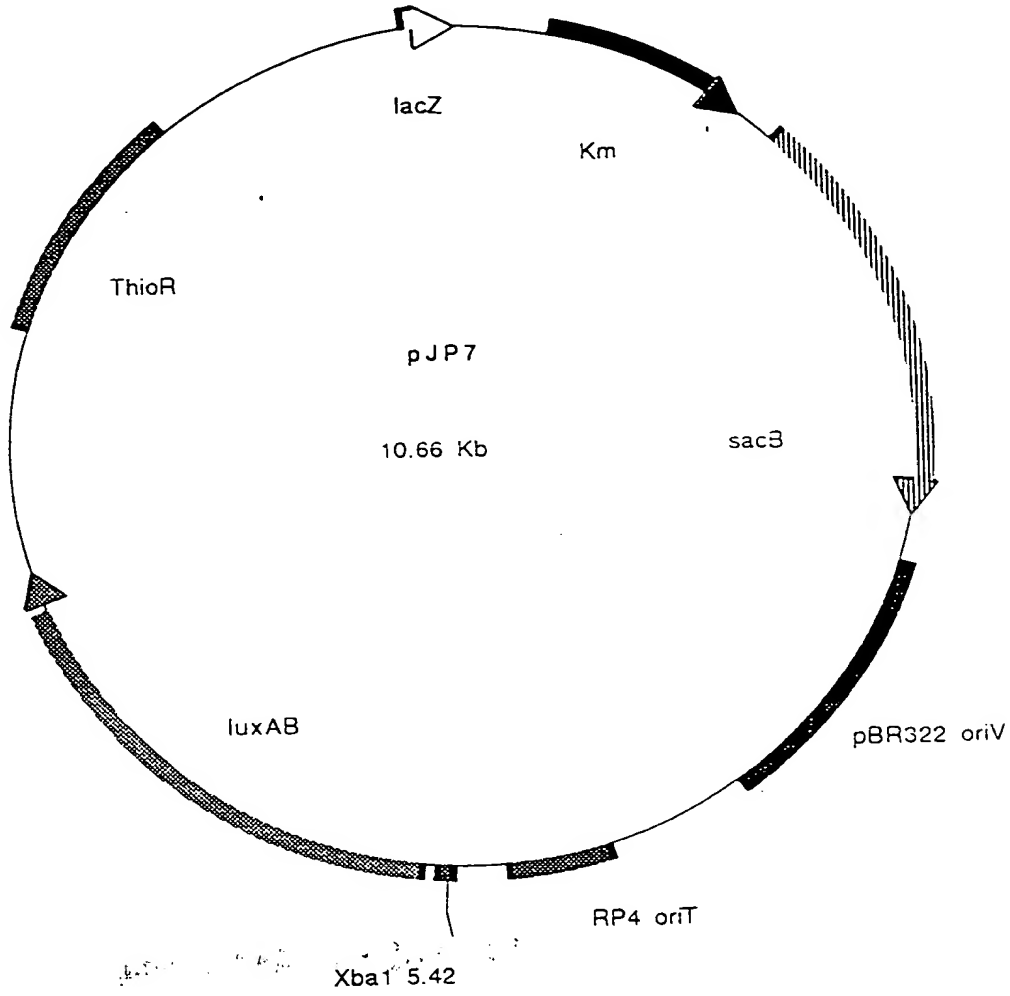
Plasmid name: pJP8
Plasmid size: 8.51 kb
Constructed by: J. Powell, H. Jacquiau & J. Archer
Construction date: 6.6.97
Comments/References: mycolic acid bacterium - Escherichia coli mobilizable cosmid vector. Carries pCY104 replion. Kanamycin resistant 15 µg/ml mycolic acid bacteria; 50 µg/ml Escherichia coli. Carries lambda cos site, RP4 oriT site. Multiple cloning site.

11/12

Figure 1

Fig. 6

Patent Specification "BACTERIAL GENE INDUCTION" Archer et al., 1997



Plasmid name: pJP7

Plasmid size: 10.66 kb

Constructed by: J. Powell & J. Archer

Construction date: 9/27/97

Comments/References: Mobilizable E.coli/Rhodococcus/Nocardia suicide/ luciferase integration vector encoding luxAB signal genes, sacB gene and thiostreppton resistance in Rhodococcus / Nocardia only up to 75 µg/ml typically 1-10 µg/ml used in selections. RP4/RK2 mobilizable. By cloning a region of homology into the region upstream of the luxAB cassette insertoin can be targeted.

12/12

~~Figure 3~~

PCT/GB9[redacted]0 — [illegible]

Mewburn Ellis

PCT/GB98/01893.

THIS PAGE BLANK (USPTO)